Intelligent Broadband Networks

Comments - Mobile Broadband Measurement

Table of Contents

Mobile Data Background and Architecture

Note: This document focuses on currently deployed 3GPP (GSM) – based networks. 3GPP2 (CDMA) and '4G' (LTE, Wimax) are not covered but are very similar in nature. The key difference to an LTE network is that there is no circuit-switched portion, and all voice data is converged. This makes data QoS a non-optional component.

Figure 1: *Simplified mobile network* shows a stylized 3GPP R7 network. This is characteristic of today's broadly deployed mobile data.

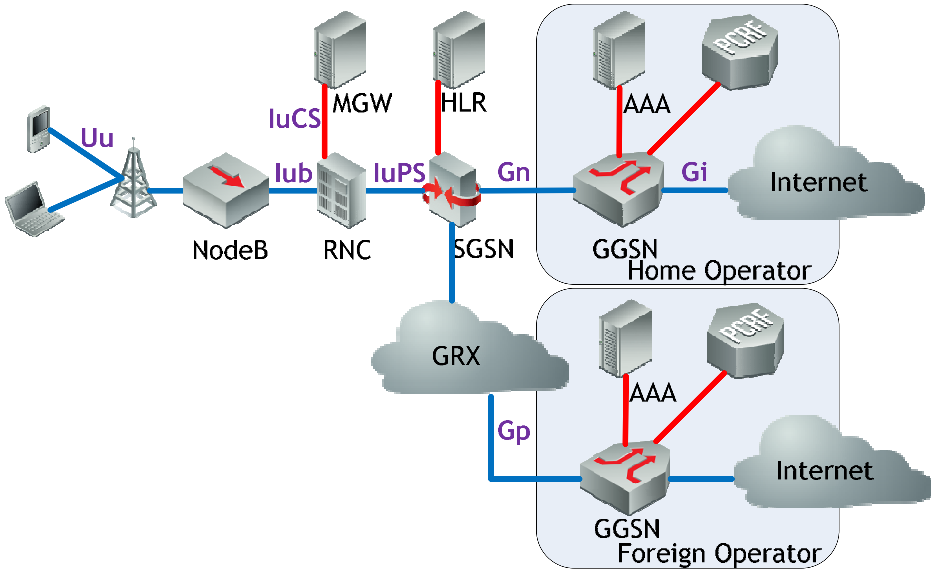

Figure 1: Simplified mobile network

The mobile network is described using a set of common acronyms and reference points. The devices present include:

- SGSN – Serving Gateway Support Node
- GGSN – Gateway GPRS Support Node
- GPRS – General Packet Radio Service
- MGW – Media Gateay (Voice to PSTN)
- HLR – Home Location Register
- AAA – Authentication, Accounting, Authorisation
- NodeB – The device which converts from fixed network to RF signaling
- NodeA – The antenna
- PCRF – Policy Charging Rating Function
- PCC – Policy Charging and Control
- RNC – Radio Network Controller
- GTP – GPRS tunneling protocol
- GRX → GPRS routing exchange (to other operators for roaming)

Reference points (named links or locations)

- Uu → over the air RF interface
- Iub → backhaul interface carrying ready-to-transmit radio packets
- IuCS → circuit switched voice path
- IuPS → packet switched backhaul data path
- Gn → GTP tunnel backhaul
- Gi → Internet connection
- Gp → GTP tunnel to other operators for roaming

There are 3 primary regions to a mobile network that must be considered when looking at quality of experience: radio access network, circuit switched (CS), and mobile packet core (PS). These are shown graphically below in Figure 2: *Mobile network subsections*. It is important that each of these be measured independently, or in such a way that the data can be broken out.

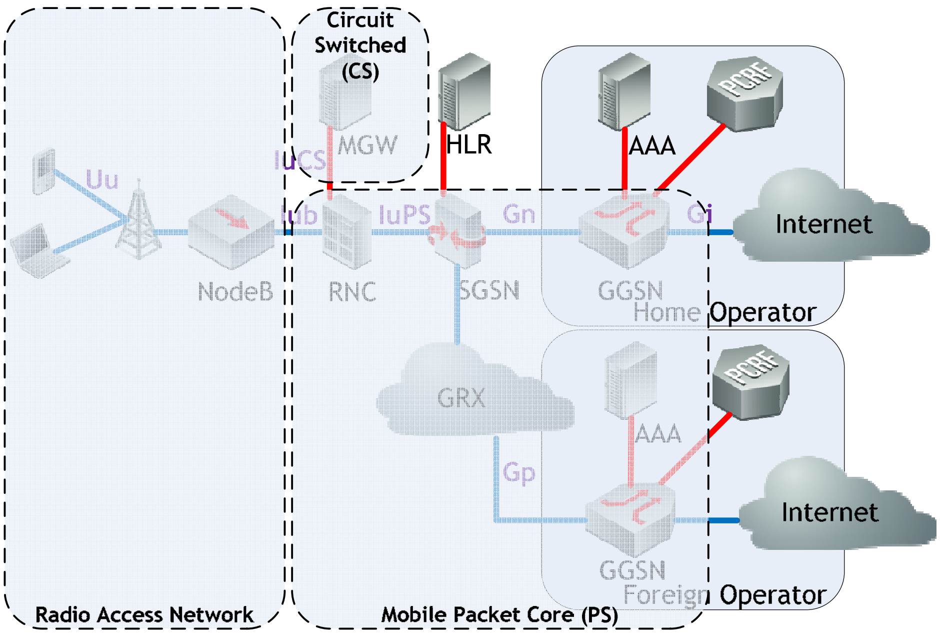

Figure 2: Mobile network subsections

Common mobile data misconceptions

- Radio usage alone is responsible for congestion

    The mobile backhaul and packet core are also oversubscribed. This ratio is getting worse as next generation networks supply greater radio bandwidth to smaller radii.

- There is a simple correlation between IP packet utilization and radio usage

    Due to packetisation and idle state management, number of devices and packet size have non-linear affects

- Data usage alone is responsible for radio usage

    Signaling, voice, SMS, MMS traffic also contribute

- Increasing bandwidth always improves experience

    Much of the user experience on a mobile data network is limited by latency

- Experience is primarily driven by applications

    In the mobile environment, experience is largely driven by the user-equipment, battery life, signal strength, consistency of experience, coverage

- All sectors have equal users

    There is a huge effect related to the number of users registered to a sector, even if they are not active. There are effects due to interference, signaling.

- The backhaul behaves like an IP-based Ethernet network

Although changing rapidly, most radio backhaul networks are based on ATM-technologies. Even with this change, there is still the use of proprietary tunneling (GTP-U) which precludes the use of IP-based congestion management (e.g. DSCP, ECN).

- The mobile data network is a broadband network with a different physical interface

  The larger latency and higher variation in performance mean certain applications do not function reliably in all circumstance on mobile. Over-subscription is much higher.

Proposed measurements

- Coverage, signal strength
- Consistency of experience
- Latency, Iups+Iub (UTRAN)
- Latency, GRX
- Latency, Packet core
- Focus measurements per device class (feature-phone, smart-phone, netbook/laptop/tablet)
- Cover case where mobile is positioned as fixed-replacement vs fixed augmentation
- Measurements should not include an 'active probe' due to battery limitation and impact on network (it alters the network under measurement)
- Measurements should consider telemetry from the user-equipment (e.g. location, signal strength)
- Measurements need to show distribution, per sector, per time, by class of device, by application class

Specific questions

1. *Best metrics to measure performance*
   a) *What performance characteristics (i.e., data throughput, signal strength, latency, other)?*

Consistency of throughput across time and location is a good measure of the overall network oversubscription. This should be done on relatively short time period (e.g. 30-seconds) rather than long durations (e.g. hour, day). This should be done by observing actual data, rather than by artificial 'speedtest' active probes, to avoid interfering with the network, and to provide a true picture of achieved experience across various handsets, locations, and times of day.

User experience does not correlate perfectly to throughput, and in a mobile environment is very much driven by latency. The majority of traffic on the network today is driven by short-lived TCP sessions (e.g. 10-50 packets). Because the TCP session is short, round-trip-latency may govern the throughput more than raw bandwidth. Mobile networks are particularly prone to latency at all levels, and measurements should be able to discern where the latency variation is introduced (UTRAN (radio), GRX (roaming exchange), Packet Core).

Other experience metrics include network equipment performance such as dropped PDP contexts (the equivalent of dropped calls) relating to too-many subscribers present. These are not directly measureable from the IP network, and can only be retrieved from network elements.

Sandvine therefore recommends measuring, in a passive fashion:

- Throughput for sentinel sites (e.g. top 100 web sites) and applications (e.g. streaming, web, email)
- Latency round-trip-time
- Network-element performance metrics of peak-concurrent PDP contexts, dropped contexts

Each of these measurements should be collected in such a fashion that it can be broken down by:

- Sector or Node-B (user-equipment geolocation)
- Subscriber tier of service
- User-equipment manufacturer and model
- Time of day
    - *b) What parts of the network should be measured?*

Sandvine recommends measuring the UTRAN (IP data within radio network), GRX (roaming IP data exchange), and the packet core, as shown in Figure 2: Mobile network subsections. Measurements should be done so as to allow presenting these independently.

c) Should measurement processes and standards be different than for fixed line?

In fixed, it is typical to ignore effects of the user-equipment. In a wireless environment, this is not a reasonable assumption. Measurements need to be able to sub-segment by manufacturer.

In a fixed environment, the majority of the latency is in the public internet, and outside the control of the access provider. In a wireless environment this is not true, and latency must be measured.

Typical fixed latencies might be < 50ms. In a wireless environment these are usually about 3-5x higher, so the threshold will be different.

In a fixed environment, a user always traverses the same path to get to the same destination. In a mobile environment the path is a function of their current location (e.g. roaming), and thus current location must be taken into account.

In a fixed environment packet loss is rare and is typically due to congestion. In a mobile environment there are many other causes for packet loss, and it cannot be considered in isolation.

2. Data collection to measure mobile broadband performance and coverage

a) What are the best available measurement tools today?

There are three types of measurement methodology:

- Active probes (which create test traffic)
- Passive probes (which measure existing traffic)
- Network element performance metrics (e.g. dropped PDP contexts)

When possible, passive probes give the most representative data. This allows measuring for each user-equipment manufacturer, for each location, for each application, exactly as the end user would use it, without creating additional network load or battery issues for the mobile devices.

Active probes give very repeatable data for bisecting problems to narrow them down, and should be used for diagnosing problems identified by passive measurement techniques.

Passive probes may be purpose-built, or may be embedded into pre-existing network elements (e.g. charging gateways, 3GPP policy charging & control enforcement functions, etc.). Passive probes typically give an end-to-end view based on where they are located.

b) Any useful current data sets available?

c) Is there existing technology today that can measure actual end-user experience? Where would these measurements take place (Device? Network?)

End user experience is a complex problem. Measurements which approximate it include:

- HTTP mean-time-to-page-load
- DNS latency
- VoIP mean opinion score (MOS)

- Consistency of throughput
- Access-round-trip latency
- Packet loss
- Minutes of streaming watched without rebuffering pauses

However, end-user experience is a subjective measure as each user may value some applications more than others.

All of the above user-experience metrics can be measured passively in the network. Active measurements on the user-equipment may exchange information with their services using techniques such as SIP-signaled RTCP-XR (extended reporting).

3. *Can user-generated data be useful to measure mobile broadband performance and coverage?*
    a) *What exists today to allow for device level data collection on broadband performance and coverage? What type of data could be collected by software on the device? What other data points would be valuable to collect (e.g., location, tower ID, handset type)?*

Measurements can be made in the network that collect performance per device, serving-area-identifier (SAI), user-equipment-type, etc.

Doing measurements on the device themselves leads to several problems:

- Inconsistent implementations by manufacturer
- Data load on network to do the measurement and report the results
- Cost to consumer for additional data load
- Device battery life and memory are limited.

For these reasons Sandvine does not suggest using the device for data measurements.

The device **is** an effective location to measure radio signal quality, and this may be periodically exchanged to the network.

b) *What impact does the type of device (smartphones, feature phones, laptop, wireless modem) have on end-user experience and network performance?*

The device type greatly affects user experience. Faster processors give snappier graphics experience. Larger screens are easier to read. These may in turn be misinterpreted by the consumer as a better 'network' performance since the overall performance is better.

Additionally, different application loads are present on a feature phone vs smart phone vs full computer. As a consequence, user expectations are different and the way they 'measure' experience varies. A feature phone uses the web through WAP, and has a high-latency, low-graphic experience. Very little in the network will improve this experience. A smart phone is optimized for email and mobile-web (a subset of the full www sites), and user experience is largely driven by email delivery times, ease of use of keyboard and screen, integration to mail systems, etc. A full computer with wireless modem will likely be viewed with the expectations of a fixed-broadband experience, involving multi-tasking, full web site usage, downloading files, etc. As a consequence, it is key to break measurements out by class of device.

Network performance too is driven by type of device. Feature phones typically connect/disconnect from the network for each data usage, creating a higher load on network systems such as AAA, GGSN, etc. as PDP contexts are created and destroyed per usage (although the overall amount of usage is less). They also place a background load on each mobile sector in terms of devices registered. The largest per-device usage comes from laptops & netbooks, and these tend to be connected in roughly 20-minute spurts of heavy usage. Smart phones tend to always stay connected and ramp usage up and down more frequently.

As smart phones and full computers blur in capabilities, and as in-between devices like tablets become commercially accepted, the per-device bandwidth utilization goes up, and the user expectation goes up.

c) *How could measurement methodology account for performance variability due to:*

   i. *location (basement, above ground) or movement (i.e., on train) of user?*

Sandvine suggests ignoring this affect, and assuming it cancels out across the network as a whole.

   ii. *Differences in location determination methods (i.e. GPS) across handsets and providers*

Sandvine recommends using the SAI (service area identifier) or CGI (cell gateway identifier), which is the mobile sector or group of sectors the device attaches to, rather than GPS location. This would be fixed, and is common across the entire industry.

   iii. *Buildings, topography, weather, continued network buildouts, other variables*

Sandvine recommends looking not at averages, but at distributions. Focus on the top-n worst sites, and this problem can be ignored. Using statistical process control, first understand the distribution, then act to shrink the variation, then act to correct the mean.

d) *d. Can sampling correct for the variables in c?*

Sandvine suggests ignoring those variables. This will lead to noise in the overall results, but it will not bias them.

e) *e. Can we measure with only minimal impact on network congestion etc.?*

Sandvine recommends using passive probes to avoid congestion, and only use active probes to diagnose known or suspected problems.

4. What are the benefits and costs of measurement for providers, regulators, customers and others? Are there any legal, security, privacy issues with collection device level data?

5. What are the best methods to publish and communicate measurement results?

A given service provider can aggregate data to anonymize individual consumers. A trusted third party is needed to normalize and amalgamate this data across carriers before publishing.

Publications should always give distributions (e.g. histograms) rather than mean or median measurements.

6. What measurements are typically performed by service providers today to track performance and service availability? What is the state of the art? How is data on coverage and performance collected, verified and displayed? What technologies support the measurements? Are there any voluntary industry standards in the area and could they be improved?